HOMEWORK 4
Mathieu Gaillard
February 10, 2019

Problem 0: Homework checklist

I mainly used the class material, Google, Wikipedia and my previous knowledge.

Problem 1: Log-barrier terms

The basis of a class of methods known as interior point methods is that we can handle non-negativity constraints such as $\mathbf{x} \geq 0$ by solving a sequence of unconstrained problems where we add the function $b(\mathbf{x}; \mu) = -\mu \sum_i \log(x_i)$ to the objective. Thus, we convert

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \geq 0 \end{array}$$

into

$$\text{minimize} \quad f(\mathbf{x}) + b(\mathbf{x}; \mu).$$

Question 1:

Explain why this idea could work. (Hint: there's a very useful picture you should probably show here!)

This idea could work because the objective function tends to infinity as x approaches the infeasible region because $\lim_{x_i \to 0^+} -\log(x_i) = +\infty$. So, with an initial $x_0 > 0$, while minimizing, the algorithm cannot choose to go to a negative value because the objective function becomes exponentially big. See the Figure 1. When $\mu \to 0^+$, the solution to the new problem tends to the solution of the initial problem. Also, a nice property is that the log function is convex so it does not breaks the convexity of the objective function.

Question 2:

Write a matrix expression for the gradient and Hessian of $f(\mathbf{x}) + b(\mathbf{x}; \mu)$ in terms of the gradient vector $g(\mathbf{x})$ and the Hessian matrix $\boldsymbol{H}(\mathbf{x})$ of f.

$$h_\mu(\mathbf{x}) = f(\mathbf{x}) + b(\mathbf{x}; \mu)$$

$$\frac{\partial h_\mu(\mathbf{x})}{\partial x_i} = \frac{\partial f(\mathbf{x})}{\partial x_i} - \mu \frac{1}{x_i}$$

$$\frac{\partial^2 h_\mu(\mathbf{x})}{\partial x_i^2} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i^2} + \mu \frac{1}{x_i^2}$$

$$\forall i \neq j, \frac{\partial^2 h_\mu(\mathbf{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\mathbf{x})}{\partial x_i \partial x_j}$$

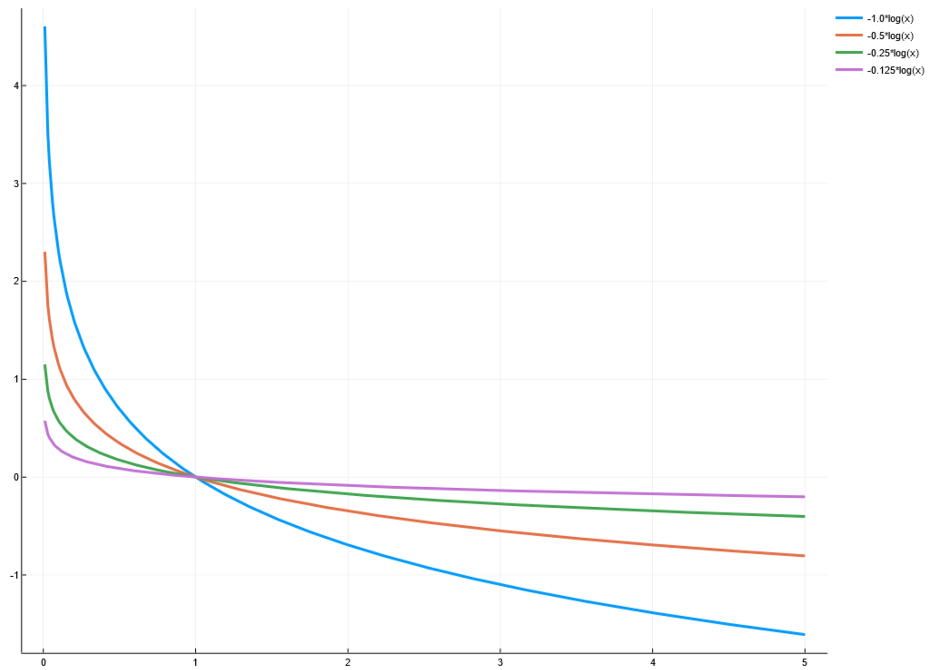

Figure 1: Plot of $-\mu \log(x)$ for $\mu = \{0.125, 0.25, 0.5, 1.0\}$.

So, in terms of the gradient vector $g(\mathbf{x})$ and the Hessian matrix $\boldsymbol{H}(\mathbf{x})$ of f:

$$\nabla h_\mu(\mathbf{x}) = \nabla f(\mathbf{x}) - \mu \left[\frac{1}{x_i}\right]_i$$

$$\nabla^2 h_\mu(\mathbf{x}) = \boldsymbol{H}(\mathbf{x}) + \mu \operatorname{diag}\left(\frac{1}{x_i^2}\right)_i$$

Problem 2: Inequality constraints

Draw a picture of the feasible region for the constraints:

$$\begin{bmatrix} 1 - x_1 - x_2 \\ 1 - x_1 + x_2 \\ 1 + x_1 - x_2 \\ 1 + x_1 + x_2 \end{bmatrix} \geq 0.$$

The feasible region is a square centered at the origin and rotated by 45 degrees, See Figure 2.

Problem 3: Necessary and sufficient conditions

Let $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \boldsymbol{Q}\mathbf{x} - \mathbf{x}^T \mathbf{c}$.

Question 1 and 2:

Write down the necessary and sufficient conditions for the problem:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \geq 0 \end{array}.$$

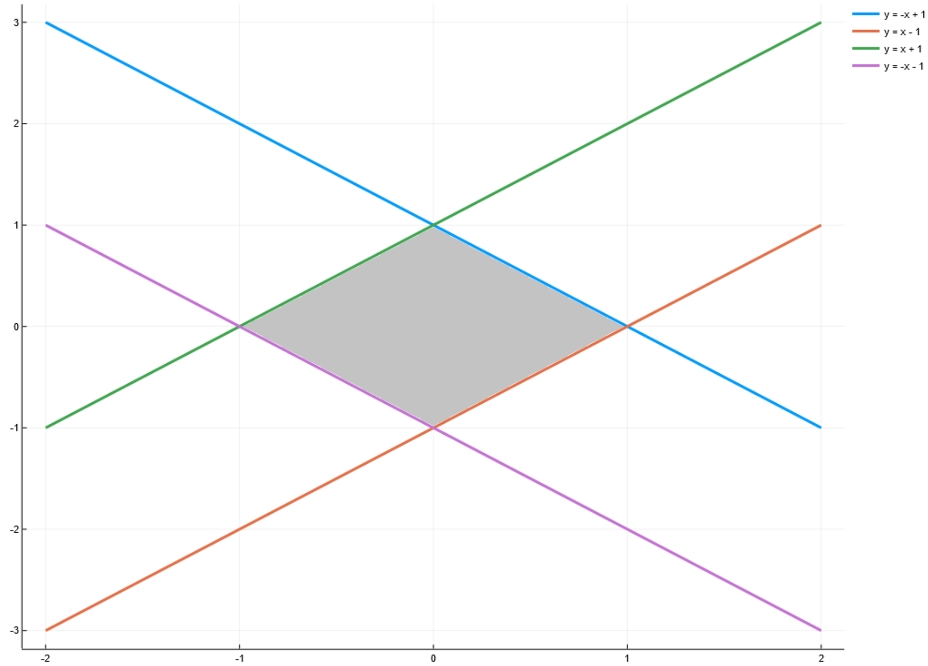

Figure 2: The feasible region is in gray.

If $\mathbf{x}^* \in \mathcal{R}^n$ is a local minimizer, then the necessary conditions are:

$$\exists \lambda^* \in \mathcal{R}^m, \mathbf{x}^* \geq 0$$
$$g(\mathbf{x}^*) = \lambda^*$$
$$\lambda^* \geq 0$$
$$\boldsymbol{Z}^T \boldsymbol{H}(\mathbf{x}^*) \boldsymbol{Z} \succeq 0$$
$$(\lambda^*)^T \mathbf{x}^* = 0$$

If $\mathbf{x}^* \in \mathcal{R}^n$ is a local minimizer, then the sufficient conditions are:

$$\exists \lambda^* \in \mathcal{R}^m, \mathbf{x}^* \geq 0$$
$$g(\mathbf{x}^*) = \lambda^*$$
$$\lambda^* \geq 0$$
$$\boldsymbol{Z}^T \boldsymbol{H}(\mathbf{x}^*) \boldsymbol{Z} \succ 0$$
$$(\lambda^*)^T \mathbf{x}^* = 0$$

$\boldsymbol{Z}$ is a basis of the feasible region, which can change according to active constraints. When no constraint is active, $\boldsymbol{Z} = I$. When the i^{th} constraint is active, $\mathbf{e}_i \in \boldsymbol{Z}$. When all constraints are active, the feasible region is just a point and $\boldsymbol{Z} = 0$.

Question 3:

Consider the two-dimensional case with

$$\boldsymbol{Q} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{c} = \begin{bmatrix} 0 \\ -1.5 \end{bmatrix}.$$

Determine the solution to this problem by any means you can, and justify your work.

With a MATLAB code we plot the function and run a constrained optimization algorithm that tells us that the minimum is most probably the origin. We also notice that:

$$\forall(\mathbf{x}) \in \mathcal{R}^{+^2}, f(\mathbf{x}) \geq 0 \quad \text{and} \quad f\left(\begin{bmatrix}0\\0\end{bmatrix}\right) = 0$$

Which means by definition that $[0;0]$ is a global minimizer on the region where $\mathbf{x} \geq 0$.

Question 4:

Produce a Julia or hand illustration of the solution showing the function contours, and gradient. What are the active constraints at the solution? What is the value of λ in $\boldsymbol{A}^T\lambda = \mathbf{g}$?

The contour and gradient plot has been generated with MATLAB. See the code in the following listing, and the result in Figure 3. All the constraints are active at the solution. The value of λ is equal to the gradient at the solution.

$$\nabla f = \boldsymbol{Q}^T\mathbf{x} - \mathbf{c}$$

So, $\lambda = -\mathbf{c} = \begin{bmatrix}0\\1.5\end{bmatrix}$.

```
function [s] = problem3()
    Q = [1 2;2 1];
    c = [0; -1.5];

    fun = @(x)general_function(Q, c, x);
    options = optimoptions('fmincon', ...
        'Algorithm', 'trust-region-reflective', ...
        'SpecifyObjectiveGradient', true, ...
        'HessianFcn', 'objective');

    % No linear constraints
    A = [];
    b = [];
    Aeq = [];
    beq = [];

    % No non linear constraint
    nonlcon = [];

    % Constraints on x
    lb = [0,0];
    ub = [];

    x0 = [1; 1];
    s = fmincon(fun, x0, A, b, Aeq, beq, lb, ub, nonlcon, options);

    % Plot the surface
    [X, Y] = meshgrid(0:0.1:2);
    Z = zeros(length(X), length(Y));
    DX = zeros(length(X), length(Y));
    DY = zeros(length(X), length(Y));
    for i = 1:length(X)
        for j = 1:length(Y)
            [f, g] = fun([X(i, j); Y(i, j)]);

            Z(i, j) = f;
            DX(i, j) = g(1);
            DY(i, j) = g(2);
        end
    end
    figure
    contour(X, Y, Z);
    hold on
    quiver(X, Y, DX, DY);
    hold off
```

```
end

function [f, g, H] = general_function(Q, c, x)
    % Calculate objective f
    f = 0.5*x'*Q*x - x'*c;

    if nargout > 1 % gradient required
        g = Q*x - c;

        if nargout > 2 % Hessian required
            H = Q;
        end
    end
end
```

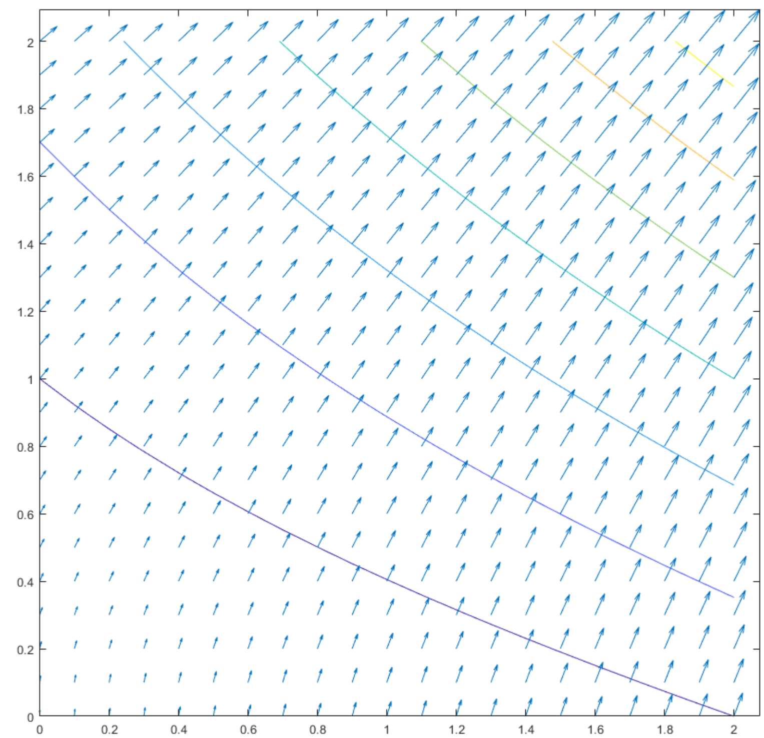

Figure 3: Contour plot of f

Problem 4: Constraints can make a non-smooth problem smooth

Show that

$$\text{minimize} \quad \sum_i \max(\mathbf{a}_i^T \mathbf{x} - \mathbf{b}_i, 0)$$

can be reformulated as a constrained optimization problem with a continuously differentiable objective function and both linear equality and inequality constraints.

The objective function is not smooth because there is a discontinuity when $\mathbf{a}_i^T \mathbf{x} - b_i$ goes from positive to negative. We could add a constraint that prevents this term from being negative. Although the problem would not be strictly equivalent, a solution of the new problem would also solve the initial problem. Note that:

$$\forall i \in \{1, 2 \ldots, n\}, \mathbf{a}_i^T \mathbf{x} - \mathbf{b}_i \geq \implies \max(\mathbf{a}_i^T \mathbf{x} - \mathbf{b}_i, 0) = \mathbf{a}_i^T \mathbf{x} - \mathbf{b}_i$$

Therefore, the new problem would be:

$$\begin{array}{ll} \underset{\mathbf{c}}{\text{minimize}} & \mathbf{e}^T \mathbf{c} \\ \text{subject to} & \mathbf{c} = \mathbf{a}_i^T \mathbf{x} - \mathbf{b}_i \quad \text{and} \quad \mathbf{c} \geq 0 \end{array}$$

The new objective function $\mathbf{e}^T \mathbf{c}$ is simply the sum of all components in $\mathbf{c}$. It is obviously continuously differentiable.